**AI 角色扮演网页｜PRD 合订本**

版本：v1.0 日期：2025-09-29 编写：产品/研发/QA

**目录（手动）**

1. 第 1 章｜用户细分（Segmentation）

2. 第 2 章｜痛点分析（Problem Statement）

3. 第 3 章｜用户画像与用户旅程地图

4. 第 4 章｜竞品分析报告

5. 第 5 章｜需求优先级与评分（MoSCoW / 90分制）

6. 第 6 章｜设计功能清单表（MVP 功能）

7. 第 7 章｜未来功能清单表（P1/P2 扩展）

8. 第 8 章｜MVP 清单与验收用例

9. 第 9 章｜ADR-001 模型选型

10. 第 10 章｜功能清单表（汇编版）

**第 1 章｜用户细分（Segmentation）**

• 舆情/民调：AP–NORC、Pew 等对生成式 AI 使用频率与分布的年度调查。

AI 角色扮演语音聊天网页  
用户细分（Segmentation）

版本：v1.1   
适用范围：面向中国与全球市场的产品规划、增长与数据分析

0. 目标与适用

本文件用于指导“AI 角色扮演语音聊天网页（MVP）”的人群策略、功能优先级、投放与指标设计。聚焦两大主力人群：好奇探索型年轻人与情感陪伴/心灵疗愈型（18+）。

1. 细分总览（中国 & 全球）

P0｜好奇探索型年轻人（18–34）

• 为什么：角色/陪伴型平台用户以年轻人占主流；海外平台数据显示 18–34 岁占比高。

• 中国侧证：AI 原生应用 MAU 快速增长（豆包/Kimi/文小言等），反映年轻用户对新奇玩法与探索偏好强。

• 全球侧证：<30 岁人群在生成式 AI/聊天机器人使用频率更高。

• 产品定位：一键开聊/随机探索、角色榜单、剧情分支、可分享聊天海报。

• 关键信号：高频切换角色/主题、剧情分支点击、长会话、晚间活跃。

• 成功指标：日会话时长、次日&7日留存、角色复访率、分享率。

P0｜情感陪伴 / 心灵疗愈型（18+）

• 为什么：全球“社会联结/孤独”成为健康议题；大量用户把聊天机器人作为情感陪伴对象（非刻意）。

• 中国侧证：本土大模型与应用渗透广（如文心一言用户规模可观），长期陪伴型对话的供需基础具备。

• 产品定位：温和共情的“陪伴模式”、语音对话、心情卡/周回顾、正向引导、敏感词触发时的求助资源引导。

• 关键信号：单角色粘性高、深夜/夜间活跃、情绪/日记功能使用、长时连续对话。

• 成功指标：7/30日留存、复购率、满意度、负向事件（困扰/不适）申诉率下降。

P1｜IP/二次元粉丝（18+）

• 为什么：与前两类人群高度重叠，但对“人设还原+追更剧情”付费意愿更强。

• 产品定位：角色设定卡、剧情分支、收藏/追更、角色商店。

• 关键指标：ARPPU、剧情完播率、追更率。

P1｜内容创作者/主播

• 为什么：用 AI 角色做直播/短视频，要求低延迟、可控人设与授权。

• 产品定位：低延迟文本→语音、台词板/热键、基础授权说明、OBS 辅助（后续）。

• 关键指标：端到端延迟 P95 ≤ 600ms、专业订阅次月留存 ≥ 50%。

2. 关键“证据窗”（中国 & 全球）

• 中国：AI 原生应用 MAU 高位运行（如：豆包/Kimi/文小言等头部应用的千万级月活），显示强烈的新奇/探索需求。

• 中国：本土大模型与应用（如文心一言）用户规模与企业采用率提升，支撑长期对话与陪伴类场景。

• 全球：角色/陪伴型平台的用户年龄集中在 18–34 岁，Gen Z 占比高，单次停留时长显著。

• 全球：亚太地区对 AI 的态度更积极，年轻群体使用频率与多场景渗透率更高。

• 全球：孤独/社会隔离的健康影响受到 WHO 等机构关注，为“陪伴型”动机提供宏观背景。

3. 线上识别与贴标规则（7日窗口）

• 好奇探索型：使用 ≥3 个不同角色；随机探索/发现页进入 ≥1；剧情分支 ≥2；单次会话中位数 ≥6 分钟。

• 情感陪伴型：同一角色会话占比 ≥70%；夜间活跃（21:00–02:00）占比 ≥40%；心情卡/日记 ≥2 次/周；≥15 分钟长会话占比高。

• IP/二次元粉丝：收藏/追更 ≥2；剧情分支 CTR（P50）≥50%；有角色/剧情付费 ≥1。

• 创作者/主播：TTS 时长 ≥30 分钟/周或 live\_hotkey ≥10；高并发（≥3）事件 ≥2 次。

• 冲突优先级：情感陪伴 > 创作者 > 好奇探索 > IP 粉丝（基于 LTV 与支持成本假设）。

4. “功能—内容—渠道”打法（对应细分）

A. 好奇探索型（P0）

• 功能：一键开聊 / 随机探索、角色榜单、剧情分支、聊天海报分享。

• 内容：每周主题挑战（如“跨界联动角色”）、彩蛋任务。

• 渠道：中国：抖音/小红书/B站；海外：YouTube/TikTok。

• 指标：会话时长 P50 ≥ 8 分钟、次日留存 ≥ 35%、7 日留存 ≥ 25%。

B. 情感陪伴/心灵疗愈（P0）

• 功能：温和回应节奏（复述-共情-建议）、语音对话、心情卡/周回顾、敏感词触发的求助资源引导。

• 内容：晚间“安心电台”模式、正向心理学微练习（非医疗）。

• 渠道：微信社群/播客/长帖故事征集。

• 指标：30 日留存 ≥ 20%、夜间活跃留存提升、申诉/不适事件率下降。

C. IP/二次元粉丝（P1）

• 功能：人设卡、剧情分支、追更/通知、角色商店。

• 内容：主题周、二创征集、联动短剧。

• 指标：ARPPU、剧情完播率、追更率。

D. 创作者/主播（P1）

• 功能：低延迟文本→语音、台词板/热键、授权说明、OBS 小助手（后续）。

• 指标：端到端延迟 P95 ≤ 600ms、专业订阅次月留存 ≥ 50%。

5. 合规与边界

• 用途提示：非医疗/非心理诊断；敏感词触发时提供本地求助热线/资源链接。

• 内容安全：对陪伴/浪漫交互设置边界；支持“一键清空记忆”。

6. 指标看板

• 分群看板：留存（D1/D7/D30）× 会话时长 × 付费转化。

• 成本看板：人均 TTS 时长成本 / 会话推理成本 / 角色版权成本。

• 增长看板：分享率、二创视频量、渠道获客 CAC。

附录 ：参考与数据来源（摘要）

• 中国：QuestMobile（2024/2025）关于 AI 原生应用与行业趋势报告（MAU 规模与头部应用）。

• 中国：路透社对文心一言用户规模与企业采用的报道。

• 全球：Similarweb/行业报道对角色/陪伴型平台用户年龄结构的统计。

• 全球：Ipsos《AI Monitor 2024》对中国及亚太用户态度与期待的调查。

• 全球：WHO（2025）关于孤独/社会联结的旗舰报告（健康影响与社会成本）。

• 学术：MIT 团队关于 AI 伴侣社群中“非刻意情感关系”形成的计算分析论文。

**第 2 章｜痛点分析（Problem Statement）**

缓解：多档模型路由与降级；合规自检与风格替代库；灰度发布与 A/B 测试。

**AI** 角色扮演 **-**用户细分痛点分析

一、项目背景与目标（简）

面向“AI 角色扮演 + 语音聊天”的网页应用，围绕五类核心用户：1）年轻探索者/泛娱乐IP粉丝；2）情感陪伴/疗愈人群；3）语言学习者（PTE/IELTS口语）；4）知识提升/苏格拉底式学习者；5）内容创作者/主播/运营。目标是让用户在首 3 分钟感知到“好玩/有用/被理解”，并形成 7 日复访与付费转化。

二、细分用户痛点分析

1) 年轻探索者 / 泛娱乐IP粉丝

JTBD（Jobs To Be Done）

核心 Job：在碎片时间里快速发现并试玩一个“有梗且沉浸”的角色，决定要不要继续玩和分享。

功能 Job：发现→预览→一键试聊；稳定人设；顺畅低时延语音。

情感 Job：被惊喜、不尴尬；梗被接住。

社会 Job：可分享、拉新好友一起玩。

成功标准：<3 分钟“哇”感；首会话≥5 轮；次日/7 日回访。

5 Whys（围绕“首会话流失高”）

找不到想聊的角色/入口深；

推荐非兴趣/情绪驱动；

冷启动无轻量偏好采集；

担心问卷打断，且缺“一键试玩/热门片段”；

早期偏向模型接入，未做“发现—预览—试聊”一体化。

Problem Statement

对于追求新鲜感与零学习成本上手的年轻探索者，当前在角色发现、预览与极速试聊上缺乏兴趣/情绪导向与一键试玩，人设和语音链路稳定性不足，致使他们无法在 3 分钟内获得直观价值，首会话流失高、复访弱。

HMW 机会

HMW 用“兴趣/情绪引导 + 15 秒语音预览 + 一键试玩”降低发现/决策成本？

HMW 以“角色卡 + 风格锚点”稳定人设与口癖？

HMW 将端到端语音往返压到 <2–3s，并可观测？

2) 情感陪伴 / 疗愈人群

JTBD

核心 Job：在低谷或孤独时获得不评判、可随时响应、能记住我的安全空间，缓解压力并恢复稳定。

成功标准：稳定可用（尤其夜间）、温和共情、轻记忆可控、隐私明确可撤回。

5 Whys（围绕“陪伴感不足”）

回复模板化、AI感强；

长期记忆弱，情境/偏好难延续；

用户不敢开记忆，隐私不透明；

设置深、文案抽象、缺“可视化与一键清除”；

安全合规与体验未具体到 UI 与技术方案（本地/云可选）。

Problem Statement

对于需要匿名且被理解的陪伴用户，系统在共情质量、可控记忆与隐私透明上不足，关键时刻难以建立信任与慰藉，尝试后难以长期留存或转化。

HMW 机会

HMW 将隐私做“可视化栈”（本地/云/留存时长/一键清除/撤回）？

HMW 训练“温柔不空洞”的共情意图与边界？

HMW 提供“夜间安心模式 + 危机资源转介”？

（边界说明：非医疗/心理诊疗；需安全边界与转介。）

3) 语言学习者（PTE/IELTS 口语）

JTBD

核心 Job：以短时高频演练，获得考纲映射 + 可执行即时纠错，短期内看到量化提升。

成功标准：维度对齐（fluency/grammar/pronunciation），词-句-音素三级纠错，学习曲线可视化。

5 Whys（围绕“练了不涨分”）

练习不对考纲与评分维度；

题型/维度未绑定脚本与反馈；

产品偏闲聊，未任务化；

缺基准语料与评分标尺；

未做“最小切片”优先（如 RS/WFD）。

Problem Statement

面向提分目标的学习者，当前缺考纲强绑定的任务化练习与可执行纠错，练习时长难转化为稳定分数增益，影响留存和付费。

HMW 机会

HMW 先聚焦 RS/WFD 两大高权重任务，提供逐题维度评分？

HMW 输出“音素级纠错卡 + 下次练习清单 + 进度曲线”？

HMW 保障低时延以维持语流节奏？

4) 知识提升 / 苏格拉底式学习者

JTBD

核心 Job：在复杂主题上被引导提问、分层推进，并沉淀结构化产出（要点/术语/行动计划），避免“看懂→忘记”。

成功标准：反诘树/难度分层；自动笔记/错题；续学。

5 Whys（围绕“学完就忘”）

输出弱，只输入不产出；

缺自动笔记/复盘；

对话产品未将“产出物”设为交付；

未定义“学习闭环”；

无面向学习成效的里程碑与 KPI。

Problem Statement

对于追求迁移与长期积累的学习者，当前缺反诘式引导、难度编排与可复盘产出，进展难延续，难转化为实际能力。

HMW 机会

HMW 用“反诘树 + 示例→抽象→迁移”的脚手架？

HMW 自动沉淀“要点/术语/错题/待验证”，生成“下次计划”？

HMW 设计“课堂/研讨模式”，控制认知负荷？

5) 内容创作者 / 主播 / 运营

JTBD

核心 Job：持续产出批量、风格一致、可多人设联动的脚本/台词，提升转化并控风险。

成功标准：模板批量、多角色协同、风格锁定、A/B 钩子、合规提示。

5 Whys（围绕“效率低且风格不稳”）

单次生成不可批量；

无模板 + 变量机制；

风格不稳，缺“口癖锚点/语气参数”；

生成提示未参数化人物特征；

未把“风格一致”作为一等公民指标。

Problem Statement

面向高产的创作者，产品缺批量化与风格锁定，合规边界模糊，导致复用率低、生产链断点多、平台风险高。

HMW 机会

HMW 提供“剧情模板库 + 变量占位 + 批量出片”？

HMW 用“角色卡 + 口癖/语速/修辞锚点”固化风格？

HMW 嵌入“合规自检清单与授权提示”？

三、跨细分的共性痛点（平台层）

● 低时延与稳定性不足

- 痛点：语音端到端往返 >2–3s 明显破坏沉浸与流利度；高峰期抖动与掉线。

- HMW：多地域边缘接入；流式 ASR/TTS 并行；断点续传；QoS 可观测（P95/P99）。

● 人设一致性与上下文连续性弱

- 痛点：长对话人格漂移、遗忘历史；多轮后“你是谁”失真。

- HMW：角色卡 + 风格锚点 + 约束解码；轻/重记忆分层；重要事实 pin 住。

● 发现与匹配体验弱（冷启动）

- 痛点：纯列表/搜索难命中兴趣/情绪/目标；无一键试玩与片段预览。

- HMW：兴趣/情绪引导；15 秒高光片段；“今日三个你可能喜欢”。

● 隐私与信任不透明

- 痛点：语音/情感数据的存储位置、时长、用途不清晰，用户不敢开记忆。

- HMW：隐私可视化面板（本地/云/留存/撤回/导出）；默认最小化采集。

● 价值与付费点不清晰

- 痛点：试用期看不见“硬价值”（提分/产出/情绪缓解），过早付费墙。

- HMW：3 分钟可感知价值路径；“新手任务/成就/里程碑”；按功能解锁而非时长。

● 多口音/多语种/可访问性适配不足

- 痛点：非标口音 ASR 误识；听障/视障辅助缺失。

- HMW：口音鲁棒 ASR、字幕与速率控制、读屏与键盘操作友好。

● 合规与 IP 边界不清晰

- 痛点：二创/商用风险；未明确授权范围。

- HMW：“合规自检 + 授权提示 + 替代风格库”。

● 成本与性能权衡

- 痛点：高并发下推理成本高、体验波动。

- HMW：多档模型路由（预览/标准/进阶）；缓存热门角色；离线 TTS 回退。

四、平台级汇总 Problem Statement（面向 MVP）

面向探索、陪伴、学习与创作等多场景，产品在低时延稳定、人设一致、冷启动发现、隐私透明、目标导向价值表达上存在系统性短板，导致首 3 分钟价值感不足与 7 日留存/转化受限。MVP 须以“可观测的低时延语音链路 + 人设锁定与轻记忆 + 兴趣/情绪引导的一键试玩 + 隐私可视化 + 目标导向交付物（提分/脚本/笔记）”作为核心设计约束。

五、平台级机会清单（Backlog 摘要）

语音链路：ASR/TTS 流式并行；P95 往返 < 2–3s；高峰自动降级策略。

角色一致：角色卡/口癖锚点/边界提示；重要事实 Pin；多轮一致性评估。

冷启动：兴趣/情绪引导、15 秒片段预览、一键试玩、三选一极速开聊。

隐私与记忆：本地/云可选、存留时长、撤回/导出、默认最少记忆。

目标交付物：创作者（模板+变量批量、多角色联动、A/B 钩子、合规自检）；知识提升（反诘树、自动笔记、续学计划）；陪伴（夜间安心、温和回复、危机资源）。

|  |  |  |
| --- | --- | --- |
| 维度 | 指标 | 目标 / 阈值 |
| 体验 | 语音 E2E 往返 P95 | < 2.5s |
| 体验 | 会话中断率 | < 2% |
| 留存 | 首会话 ≥5 轮占比 | ≥ 60% |
| 留存 | D1 / D7 留存 | ≥ 35% / ≥ 18% |
| 转化 | 3 分钟内首次价值事件达成 | ≥ 65% |
| 一致性 | 人设漂移投诉率（/千会话） | < 1% |
| 一致性 | 关键事实一致得分 | ≥ 0.9 |
| 隐私信任 | 开启可控记忆占比 | ≥ 40% |
| 隐私信任 | 隐私设置访问转化率 | ≥ 70% |
| 学习成效 | 两周内模拟分提升（任务化场景） | ≥ 5–10 分 |
| 创作效率 | 批量脚本平均产出时长下降 | ≥ 40% |

六、验证指标（KPI 建议）

七、关键风险与假设

假设：3 分钟可感知价值能显著提升 D1/D7；用户愿意接受轻量偏好采集。

风险：高并发下时延与稳定不可控；隐私/合规规范变化；IP 争议。

**第 3 章｜用户画像与用户旅程地图**

访谈：口吻一致性/边界可理解性、要点卡是否有学习价值。

AI 角色扮演网页 — 用户画像（Persona）& 用户旅程地图（User Journey Map）

日期：2025-09-24  
版本：v1.0（面向 MVP：先“文本→文本”，后续接入 ASR/TTS 与情绪识别）

0. 背景与范围（Scope）

产品目标：

让用户搜索并与喜欢的“角色”（如：哈利波特、苏格拉底等）进行对话（MVP 先文本→文本），逐步升级为语音聊天（ASR/TTS）与基础情绪识别（情感标签驱动回复风格）。

**MVP** 范围：

- 角色设定卡（性格/口吻/知识边界）

文本对话（启发式追问、口语纠错要点、结构化笔记 3 技能骨架可复用）

- 历史会话/收藏夹

- 语音（ASR/TTS）、情绪识别（后续迭代）

关键假设：

- 核心早期用户来自“英语学习/教育场景/内容创作者”三类；

角色设定的“稳定口吻 + 可解释边界”决定满意度；

- 低延迟与稳定性>炫酷功能。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Segment | 画像关键词 | 使用动机 | 场景 | 优先级 |
| 学习者（英语/哲学/历史） | 大学生/考生/自学者 | 练口语/泛读、带趣味的知识探索 | 与角色讨论剧情/观点；做口语纠错与关键词笔记 | ★★★★☆（主攻） |
| 教育工作者 | 中学/大学教师、培训讲师 | 课堂互动、苏格拉底式提问、备课引子 | 课堂演示、小组讨论、家庭作业提示 | ★★★★☆ |
| 内容创作者 | B 站/小红书/抖音 UP | 趣味脚本、角色同人桥段、互动素材 | 生成对话脚本、分享卡片 | ★★★☆☆ |
| 泛娱乐粉丝 | IP 粉、二次元 | 代入感与陪伴、角色延展 | 日常闲聊、剧情脑洞 | ★★☆☆☆ |

1. 目标用户分层（Segments & Priority）

**Primary Persona**：学习者（英语学习 + 知识探索）

2. Persona（3 张）

Persona A｜林晨（23）— 英语学习者 & 影迷（主画像）

基本信息：大三在读，iPhone + 笔记本；通勤碎片时间多；YouTube/B 站重度用户。

动机**/**目标：

用喜欢的角色“陪练英语”，提升关键词捕捉与表达流利度；

轻负担获取剧情/知识；对“可保存要点卡片”有刚需。

行为特征：

1～3 分钟短对话为主；喜欢自动要点提炼与纠错高亮；

接受轻量注册，但不愿复杂设置。

痛点：

生词多、听力抓不住关键词；

平台人设不稳、回答“出戏”；

延迟与中断严重影响沉浸。

成功判据（**KPI**）：

首次 3 分钟内连贯 5 回合；

收藏/要点卡 ≥1；次日复访。

引述： > “我需求学术正确，并且希望像角色本人那样陪我练习，还能帮我标关键词与纠错。”

Persona B｜周某（29）— 初中英语老师 & 教育博主

动机**/**目标：

课堂用“角色引子”带讨论；

生成苏格拉底式问题串、可投屏的讨论要点卡。

行为特征：

关注内容可控性**/**可解释性与学生安全；

课前 15 分钟快速备课。

痛点：

需要可审阅、可引用的回答依据；

校园网络/权限限制；

未成年合规与敏感话题过滤。

成功判据：一节课内产出 2～3 个可投屏卡片，学生参与度↑。

Persona C｜张弈（27）— 短视频创作者（科普/趣味）

动机**/**目标：

生成“角色×热点”的对话脚本；

一键导出分镜**/**字幕草案。

行为特征：

在 PC 端长时段编辑；

关注可商用授权、版权风险。

痛点：

口吻不稳、风格不统一；

导出不便、格式杂。

成功判据：单次 10～15 分钟得到可改编脚本。

3. Jobs To Be Done（主画像 A）

**When** 我想用熟悉 IP 以低心理负担练英语 **I want to** 快速开始与“角色”对话并被适度追问 **so I can** 形成可复习的关键词与句型卡。

验收标准（**AC**）：

搜索角色 < 10 秒有命中； 2) 首次响应首字节 TTFB < 2 s；

每回合提供“关键词高亮/一键加入要点卡”；

会话结束可导出/收藏； 5) 角色口吻一致，越界时给出边界提示。

4. 用户旅程地图（主画像 A｜文本→文本 MVP）

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 阶段 | 触发/场景 | 用户任务（前台） | 心智 & 情绪 | 关键痛点 | 机会点/设计要点 | 指标（MVP） | 服务蓝图（后台） |
| 发现 Discover | 社媒短视频/同学推荐 | 进入落地页、浏览角色示例 | “真的像本人吗？”🙂 | 信息噪音，价值不清晰 | 演示对话 + 关键词卡片样例 + 口吻保证 | LP→试用 CTR，跳出率 | 角色示例缓存、匿名试用令牌 |
| 进入 Onboard | 首次访问 | 选择“快速试用”或注册 | “先试试看。”🤔 | 注册阻力 | 免登录 5 回合试用；注册即同步历史 | 首回合启动率 | 限流/风控、匿名会话桶 |
| 选角 Search | 搜索“Harry Potter/苏格拉底” | 点选角色卡，读设定 | “别出戏就好。”😊 | 无检索命中/太多同名 | 角色卡含口吻样句/边界声明；相似角色推荐 | 角色命中率、点卡率 | 向量检索/别名索引、冷启动回退 |
| 首聊 First Chat | 输入第 1 句 | 获得 1st 回复 + 追问 | “像本人！别太慢。”😊 | 延迟、回答跑题 | TTFB<2s，首轮追问与关键词高亮 | RTT、满意度 CSAT | 低温响应模板、对齐提示器、安全过滤 |
| 深聊 Deepen | 连续 3–7 回合 | 追问/纠错/要点卡累积 | “被带着说下去。”😍 | 纠错过重打断流 | 轻量纠错（悬浮高亮）+ 一键加入要点卡 | 回合数、要点卡数 | 轻纠错器、术语本、节奏控制 |
| 结束 Save/Share | 会话收尾 | 保存/导出要点卡 | “下次接着聊。”🙂 | 导出/回看不便 | 收藏夹 + 导出 Markdown/图片卡 | 收藏率、导出率、D1 | 会话持久化、导出服务 |
| 复访 Retain | 第二天提醒 | 一键续聊/复盘测验 | “形成习惯。”🙂 | 无触发、忘记打开 | 次日复盘卡 + 续聊入口 | D1/D7 留存 | 定时任务、简报生成 |

阶段验收（**AT**）： - A/B：首屏“演示对话 + 要点卡”是否显著提升 LP→试用； - 技术：TTFB<2s、RTT<4s（P50），角色一致性评分≥0.85； - 体验：用户 3 分钟内≥5 回合、收集≥1 张要点卡。

5. 异常路径 / 边缘场景（Edge Cases）

无结果**/**同名多：给出近似角色与“创建自定义角色”入口；

越界**/**安全：敏感内容被拦截→以角色口吻给“边界说明 + 安全替代提问”；

网络**/**延迟：>4s 提示“正在召唤角色…”，先回短句 + 继续加载；

权限：未来语音需麦克风权限预检测与模拟演示；

学习进度：断网/退出后自动恢复到上次节点。

|  |  |  |  |
| --- | --- | --- | --- |
| 目标 | 指标 | 定义 | 目标值（首月） |
| 触达/转化 | LP→试用 CTR | 访问落地页到进入试用 | ≥18% |
| 首体验 | 首回合 TTFB/RTT | 首字节 / 完整回复时延（P50） | <2s / <4s |
| 参与度 | 会话回合数 | 单会话有效回合 | ≥6 |
| 功能价值 | 要点卡使用率 | 会话含≥1 卡的比例 | ≥45% |
| 口碑 | CSAT/NPS | 5 分满意度 / 推荐度 | CSAT ≥4.3 |
| 留存 | D1/D7 | 次日/7 日回访 | D1≥25%，D7≥10% |

6. 指标与埋点（MVP）

|  |  |  |
| --- | --- | --- |
| 痛点 | 设计对策 | MVP 实现 |
| 角色“出戏” | 角色卡 + 边界声明 + 风格判定 | 设定卡模板、对齐提示器 |
| 关键词抓取难 | 高亮关键词 + 一键要点卡 | 关键词抽取、收藏夹 |
| 延迟影响沉浸 | 低温模板 + 逐字流式 | 流式输出、占位骨架屏 |
| 无法复盘 | 会话要点卡导出 | Markdown/图片导出 |
| 搜索命中差 | 别名/同义索引、相似推荐 | 向量检索 + 回退兜底 |

7. 需求→功能映射（从痛点反推 MVP）

8. 后续迭代（展望）

**ASR/TTS** 语音：噪声鲁棒、断句策略、打断重说；

情绪识别：仅作为“回复风格权重”输入，避免过拟合人格评判；

课堂模式：多用户投屏、可审阅问题串、合规开关；

创作模式：脚本导出（SRT/分镜 Markdown）。

附：可用性测试要点（招募 8–12 人）

任务：搜索角色→对话 5 回合→保存 1 张要点卡→次日复访；

指标：SUS、完成率、错误率、TTFB、主观角色一致性评分；

**第 4 章｜竞品分析报告**

- MVP 版本：提供 3–5 个预设角色（如哈利波特、苏格拉底、孔子），重点打造语音输入+输出体验。  
- 迭代方向：用户可自建角色、增加情感语音（喜怒哀乐）、打造角色社区（用户分享/交流）。

竞品分析报告：AI 角色扮演语音网页

1. 产品背景

目标：开发一款 AI 角色扮演网页，用户可搜索并选择感兴趣的角色（如哈利·波特、苏格拉底等），通过语音与角色进行实时对话，体验沉浸式交互。  
关键功能：  
- 角色库搜索：用户可挑选不同角色（文学、历史、影视）  
- 语音交互：支持语音输入（ASR）和语音输出（TTS）  
- 角色一致性：AI 保持角色设定和语气风格  
- 沉浸式体验：支持长对话、多轮记忆

2. 对比维度

- 基础功能（文本/语音输入输出、角色搜索）  
- 角色扮演能力（语气/知识一致性）  
- 语音体验（TTS 自然度、延迟）  
- 用户体验（UI/交互设计、易用性）  
- 扩展性（角色库规模、第三方接入能力）  
- 成本（API 调用费用、运营成本）

3. 竞品调研

3.1 Character.AI

- 功能：支持创建/选择角色，角色保有特定人设与对话风格。  
- 优势：角色库庞大（社区贡献）、角色一致性较强。  
- 不足：语音交互弱（主要以文本为主），延迟较高。

3.2 Inworld AI

- 功能：为游戏/虚拟世界打造 AI NPC，可接入语音输入/输出。  
- 优势：角色扮演深度高，支持情绪、个性维持。  
- 不足：偏 B 端开发工具，接入复杂，成本高。

3.3 Replika AI

- 功能：主打 AI 陪伴，可选择角色风格，支持语音聊天。  
- 优势：语音交互体验流畅，情感陪伴强。  
- 不足：角色选择有限，知识一致性较弱。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 产品 | 语音支持 | 角色一致性 | 角色库规模 | 用户体验 | 成本 | 适用场景 |
| Character.AI | 文本为主 | 强 | 海量 | 中 | 免费+订阅 | 娱乐、社区 |
| Inworld AI | 有 | 很强 | 小 | 中等 | 高 | 游戏开发、NPC |
| Replika AI | 有 | 较弱 | 有限 | 强 | 订阅 | 陪伴聊天 |

4. 横向对比表

5. 分析结论与启示

1. 语音是核心卖点：Character.AI 在语音方面不足，Inworld 偏 B 端，Replika 虽支持语音但角色局限。  
2. 角色一致性决定沉浸感：需要设计角色设定模板（Prompt Engineering），保证 AI 回答符合角色风格。  
3. 差异化竞争机会：Character.AI 强在社区，Inworld 偏游戏，Replika 偏陪伴。我们可以定位在“角色多样化 + 实时语音体验”。  
4. 成本与可落地性：三者都有短板，我们的机会是轻量、低门槛、语音沉浸式体验。

6. 建议

**第 5 章｜需求优先级与评分（MoSCoW / 90分制）（源文件未找到）**

**第 6 章｜设计功能清单表（MVP 功能）**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 模块 | 功能 | 用户故事（User Story） | 优先级 | 版本 | 依赖 | 验收标准 | 指标（度量） |
| Onboarding | 着陆页价值主张 | 作为新访客，我希望 5 秒内理解产品价值并点击开始。 | P0 | MVP | — | 首屏有标题+副标题+CTA；4G 环境 LCP ≤ 2.5s | “开始聊天”按钮 CTR |
| Onboarding | 角色发现（分类/标签） | 我希望通过分类/标签快速发现感兴趣角色。 | P0 | MVP | 角色库 | ≥ 12 个角色卡片；筛选/排序可用 | 浏览→开聊转化率 |
| Onboarding | 搜索（中英双语） | 我希望按名字/系列/特征搜索角色。 | P0 | MVP | 索引 | Top1 相关结果可见；联想 ≤ 10 条 | 搜索成功率 |
| Onboarding | 游客模式 | 我希望无需注册即可先体验聊天。 | P0 | MVP | 本地存储 | 游客会话可保存；升级账号可保留历史 | 游客→注册转化 |
| 账号 | 邮箱/手机登录 | 我希望登录后可长期保存聊天记录。 | P1 | Post-MVP | 鉴权 | 基础限流；会话保持 14 天 | 带账号 DAU |
| 角色目录 | 角色详情页 | 我希望查看设定/技能/示例对话后直接开聊。 | P0 | MVP | 提示词库 | 加载 < 2s；“开始聊天”初始化人格上下文 | 详情→开聊率 |
| 合规 | IP/合规标识 | 作为运营，我需展示授权/免责声明。 | P1 | Post-MVP | 法务 | 授权/同人标识与政策链接 | 政策点击率 |
| 聊天 | 文本→文本核心聊天 | 我希望获得贴合角色人设的一问一答。 | P0 | MVP | LLM | 平均往返 ≤ 2.5s；人设一致性 ≥ 0.8 | 响应时延 |
| 聊天 | 会话历史 | 我希望查看/重命名/置顶/删除历史会话。 | P0 | MVP | 存储 | CRUD 可用；置顶固定在列表上方 | D7 留存 |
| 安全 | 内容安全与拦截 | 我希望交互安全、违规有清晰提示。 | P0 | MVP | 安全策略 | 违规拦截与提示文案明确 | 拦截率/误杀率 |
| 技能 | 启发式追问 | 我希望在表达不清时获得 1–2 个引导问题。 | P0 | MVP | 技能引擎 | 模糊意图时自动补问 | 追问互动率 |
| 技能 | 口语纠错要点 | 我希望获得 ≤3 条精炼的发音/用词建议。 | P0 | MVP | 技能引擎 | 用户开启“Speaking Tips”即返回要点 | 功能启用率 |
| 技能 | 结构化笔记 | 我希望一键生成要点笔记与小标题。 | P0 | MVP | 技能引擎 | 一键生成可编辑笔记块 | 笔记保存率 |
| 技能 | 角色专属扩展（小剧本/线索板/反诘树） | 我希望不同角色有专属工具面板。 | P1 | Post-MVP | 技能引擎 | 每角色可配置独立组件 | 组件使用率 |
| 语音 | 文本转语音（TTS） | 我希望听到角色“声音”。 | P1 | Post-MVP | TTS | 流式播放；首包 ≤ 1.2s | 语音使用占比 |
| 语音 | 语音转文本（ASR） | 我希望直接语音对话。 | P1 | Post-MVP | ASR/VAD | 自动端点；WER ≤ 15% | 语音会话占比 |
| 语音 | 音频设置 | 我希望选择设备并调速/调音高。 | P1 | Post-MVP | WebAudio | 设备可枚举；设置可持久化 | 设置变更次数 |
| 个性化 | 记忆（用户可选） | 我希望角色记住偏好（可随时关闭）。 | P1 | Post-MVP | 隐私/存储 | 明示同意；每会话可切换 | 记忆召回准确率 |
| 分享 | 会话快照/链接分享 | 我希望分享只读聊天内容。 | P1 | Post-MVP | 永久链接 | 脱敏处理；链接打开 < 2s | 每 DAU 分享数 |
| 支付 | 额度/订阅 | 作为运营，我希望变现与配额管理。 | P2 | Later | 计费 | Stripe/支付宝沙箱通过；配额拦截生效 | ARPPU |
| 分析 | 产品基础分析 | 作为 PM，我需要漏斗与留存看板。 | P0 | MVP | 事件埋点 | 事件规范落地；DAU/WAU、D1/D7 面板 | D1/D7 |
| 后台 | 角色库管理 CMS | 作为管理员，我要创建/编辑角色、标签、技能绑定。 | P0 | MVP | 管理员鉴权 | 版本化；草稿→发布流程 | 发布前置时长 |
| 后台 | 内容审核队列 | 作为审核员，我要查看与处置被标记内容。 | P1 | Post-MVP | 审核 | 队列/分级/审计日志 | 审核 SLA |
| 帮助 | 新手引导 & FAQ | 我希望在页内获得自助说明。 | P0 | MVP | 文档 | 关键位置内嵌提示 + FAQ 搜索可用 | 帮助点击率 |
| 设置 | 语言/主题/隐私 | 我希望设定界面语言、主题和数据共享。 | P1 | Post-MVP | 偏好 | 账号/设备层持久化 | 选择退出率 |

**AI** 角色扮演网页 **—** 设计功能清单表

**第 7 章｜未来功能清单表（P1/P2 扩展）**

功能清单表（**Features List**）

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 功能ID | 功能名称 | 用户价值 | 一句话用户故事 | 主要模块/流程 | 依赖 | 优先级 | 验收标准（90分制） | 归属 & 埋点 |
| F-001 | 文本对话（非语音） | 静音/无麦环境可用；稳定入口 | 作为用户，我希望直接打字与角色对话 | 输入框 → 上下文管理 → LLM 回复 → 展示 | LLM | P0 | 首屏回显≤100ms；响应成功率≥99%；多轮稳定性≥85/90 | 前端/后端｜埋点：会话开始/结束、轮次、耗时 |
| F-002 | 多轮上下文记忆 | 避免重复提问，提升沉浸 | 我希望角色能记住我刚刚说过的事 | 短期窗口记忆 + 轻量长期记忆（用户卡片） | KV 存储 | P0 | 最近10轮关键信息召回≥90/90；人格一致性≥80/90 | 后端｜埋点：记忆命中率、重提率 |
| F-003 | 角色背景知识库（人设/RAG） | 风格一致、世界观完整 | 我希望角色永远“像他自己”说话 | 人设模板→检索→提示词编排→应答 | KB/RAG | P0 | 语气一致性≥80/90；禁忌词召回率≥95% | 产品/后端｜埋点：人设命中、冲突告警 |
| F-004 | 小剧本生成（Hook/分镜骨架） | 快速给剧情钩子，降低创作门槛 | 我希望一键得到可续写的短剧情 | 模板选择→剧情生成→续写/改写 | LLM | P0 | ≥3类模板；续写连贯度≥80/90；生成≤3s | 后端｜埋点：模板使用率、续写率、导出率 |
| F-005 | 口语纠错（文本级） | 练口语但先用文本承载 | 我希望看到读错/写错处与改法 | 错误定位→纠错建议→例句/IPA | NLP/规则 | P0 | 定位准确性≥80/90；建议可读性≥80/90 | 后端｜埋点：纠错调用数、接受率 |
| F-006 | 对话摘要/结构化笔记 | 一键沉淀知识点与待办 | 我希望把长对话整理成要点 | 会话摘要→要点提取→TODO | LLM | P0 | 三档粒度；摘要可读性≥80/90；导出成功率≥99% | 前后端｜埋点：摘要使用率、导出率 |

**P0**｜**MVP**（本期必须交付）

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 功能ID | 功能名称 | 用户价值 | 一句话用户故事 | 主要模块/流程 | 依赖 | 优先级 | 验收标准（90分制） | 归属 & 埋点 |
| F-101 | 情绪识别 & 情感回应（文本） | 被理解与被安抚 | 我希望角色能读懂我情绪并对应回应 | 文本情感分类→策略话术库→应答风格化 | 分类器 | P1 | 6类情绪识别 F1≥0.80；回应契合度≥80/90 | 后端/产品｜埋点：情绪分布、安抚成功率 |
| F-102 | 反诘树（Socratic） | 启发式学习、深度思考 | 我希望被高质量追问引导 | 问法树策略→分支追问→收束总结 | 策略引擎 | P1 | 三层连贯不兜圈；总结清晰度≥80/90 | 后端/产品｜埋点：追问层数、放弃率 |
| F-103 | 每日任务 & 打卡 | 形成习惯、提升留存 | 我希望每日收到小挑战并可打卡 | 任务派发→提醒→打卡→补记 | 定时器 | P1 | 通知抵达率≥98%；7日留存提升≥基准 X% | 前后端/运营｜埋点：完成率、连续天数 |
| F-104 | 线索板/关系图（Clue Board） | 可视化角色/事件关联 | 我希望把人物与线索拖拽整理 | 节点/边管理→画布交互→导出 | 前端图形库 | P1 | 100节点内流畅；PNG导出成功率≥99% | 前端｜埋点：节点数、导出率 |
| F-105 | 同人创作助手 | 提升二创效率与风格统一 | 我想快速写同人片段且不中断人设 | 风格固化→桥段补全→衔接检查 | LLM | P1 | 风格一致性≥75/90；桥段衔接≥80/90 | 后端/产品｜埋点：使用时长、完稿率 |
| F-106 | 可分享会话卡片 | 外部传播裂变 | 我希望把精选对话一键生成卡片 | 片段选择→排版模板→导出 | 前端渲染 | P1 | ≥3种模板；分享点击回流≥X% | 前端/增长｜埋点：分享率、回流率 |

**P1**｜增强（次迭代）

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 功能ID | 功能名称 | 用户价值 | 一句话用户故事 | 主要模块/流程 | 依赖 | 优先级 | 验收标准（90分制） | 归属 & 埋点 |
| F-201 | 多人房间同聊 | 与朋友共创/同玩 | 我希望和朋友一起与角色对话 | 房间/成员管理→消息路由→并发仲裁 | 实时信令 | P2 | 3人并发无明显延迟；状态一致性≥99% | 前后端/RTC｜埋点：房间时长、掉线率 |
| F-202 | 角色联动/混战 | IP 碰撞、趣味提升 | 我希望多个角色同台互动不串台 | 多 persona 编排→轮替/联合应答 | Prompt 编排 | P2 | 串台率≤5%；用户评分≥80/90 | 后端/产品｜埋点：联动场次、满意度 |
| F-203 | 剧情任务/解谜（活动） | 深度沉浸、可复玩 | 我希望完成有分支的任务剧情 | 任务状态机→触发条件→结局收敛 | 状态机 | P2 | ≥3条可完成支线；重复游玩率≥X% | 后端/产品/运营｜埋点：完成率、分支流向 |
| F-204 | 成就系统/积分商城 | 激励与留存 | 我希望达成成就并兑换奖励 | 勋章/积分→兑换→风控 | 账户/风控 | P2 | 防刷通过；核算准确率≥99.9% | 后端/运营｜埋点：积分消耗、留存提升 |
| F-205 | A/B 话术实验平台 | 话术持续优化 | 我希望验证不同话术的效果 | 版本管理→用户分流→统计分析 | 实验平台 | P2 | 指标差异统计显著（p<0.05） | 后端/数据｜埋点：实验曝光/转化 |
| F-206 | 社区话题/活动日历 | 运营阵地 | 我希望报名活动并查看主题内容 | CMS→话题聚合→报名/提醒 | CMS | P2 | 报名成功率≥99%；内容发布≤5min 生效 | 前端/运营｜埋点：话题互动、报名转化 |

**P2**｜拓展（社交**/**玩法）

|  |  |  |  |
| --- | --- | --- | --- |
| 里程碑 | 时间窗口 | 目标功能 | 验收口径 |
| M1 | 本期 | F-001/002/003/004/005/006 | MVP 场景完整跑通，核心指标达标 |
| M2 | +1 迭代 | F-101/102/103/104/106 | 留存与分享能力上线 |
| M3 | +2 迭代 | F-105 + P1 打磨 | 二创与可视化体验提升 |
| M4 | +3 迭代 | F-201/202/203 | 社交与活动玩法形成闭环 |

里程碑映射（简表）

**第 8 章｜MVP 清单与验收用例**

|  |  |  |  |
| --- | --- | --- | --- |
| 项目 | AI 角色扮演网页（MVP） | 版本 | v1.0（文本优先）+ v1.1（TTS 增强） |
| 日期 | 2025-09-29 | 文档类型 | MVP 清单 + 验收用例集 |
| 负责人 | 产品 / 研发 / QA 共用 | 适用范围 | 搜索→选角→开聊→历史→（可选 TTS） |

AI 角色扮演网页  
MVP 清单与验收用例

1. 版本策略（推荐）

V1.0 文本优先：先闭环文本问答、角色一致性与可观察性，降低首版交付风险。

V1.1 MVP+：在 V1.0 基础上追加 TTS 播报与失败回退，不改变核心用户旅程。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模块 | 功能点 | 版本 | 说明 | 验收口径（摘要） |
| 核心聊天 | 文本→文本一问一答；人设一致；显示关键词高亮 | V1.0 | 含启发式追问/纠错要点/要点卡三项轻技能 | P50 TTFB<2s，RTT<4s；内容不出戏 |
| 角色与模板 | 角色卡（口吻/边界/口癖锚点）；≥10 模板；新建/保存/导入导出 | V1.0 | 可复用与分享 | 导入导出后设定一致 |
| 发现与搜索 | 角色发现页（≥12 卡片）；中英搜索；联想建议 | V1.0 | 错拼/别名回退；支持自定义角色入口 | 角色详情<2s 打开 |
| 会话与记忆 | 历史持久化（重命名/置顶/删除）；短期记忆 6–8 轮 | V1.0 | 本地优先，最小化存储 | 刷新后历史存在；记忆可引用 |
| 游客试用 | 免注册 5 回合体验 | V1.0 | 转化提示登录 | 可完成 ≥5 回合 |
| 合规与安全 | 边界提示与替代问法；隐私/记忆说明；一键清空 | V1.0 | 不中断会话 | 拦截文案友好 |
| 分析与指标 | LP→试用 CTR、回合数、D1/D7、要点卡使用率 | V1.0 | 轻埋点 | 埋点事件完整 |
| TTS 播报 | 点击朗读；暂停/继续/重播；失败自动回退文本 | V1.1 | 弱网降级保底文本 | 首包≤1.2s；控制可用 |

2. MVP 功能清单

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 用例编号 | 模块 | 目标 | 前置条件 | 步骤 | 预期结果/验收标准 | 优先级 |
| A-001 | 体验启动 | 首屏加载达标且 CTA 明确 | 首访，4G/普通宽带 | 打开落地页→点击“开始体验” | LCP≤2.5s，进入试用页成功 | P0 |
| A-002 | 搜索命中 | 快速命中目标角色 | 无 | 搜索“Harry Potter/苏格拉底”→选择第一条 | Top1 相关；联想≤10 条；详情<2s 打开 | P0 |
| A-003 | 首轮响应 | 首轮回复时延与质量达标 | 无 | 输入首句→等待回复 | P50 TTFB<2s，RTT<4s；含1条启发式追问；不出戏 | P0 |
| A-004 | 连续对话 | 5 回合内体验稳定 | 无 | 连续交流 ≥5 回合 | 3 分钟内达成；无错误中断；>4s 出现占位与降级 | P0 |
| A-005 | 路径闭环 | 搜索→选角→开聊→历史可查 | 无 | 完成全链路 | 成功率≥97%；历史时间倒序 | P0 |
| B-001 | 模板应用 | 模板库可用且生效 | 模板库已发布 | 打开模板库→筛选→套用 | ≥10 模板；口吻/边界注入生效 | P0 |
| B-002 | 导入导出 | 自定义角色可复用 | 有导出的模板文件 | 新建→保存→导出→导入 | 设定一致，口吻锚点/禁忌规则保留 | P1 |
| C-001 | 历史持久化 | 刷新后历史存在 | 已有结束的会话 | 结束→刷新页面 | 历史持久化；可重命名/置顶/删除 | P0 |
| C-002 | 短期记忆 | 连续话题可被引用 | 同一主题对话 | 围绕同一事实对话≥8轮 | 6–8轮关键事实被正确引用；人格不漂移 | P0 |
| D-001 | 发现页 | 浏览/筛选顺畅 | 无 | 打开“角色发现” | ≥12 角色卡；分类/标签筛选可用；详情→开聊顺畅 | P1 |
| D-002 | 无结果回退 | 错拼/冷门有替代 | 无 | 输入生僻别名/错拼 | 给出相似推荐与“创建自定义角色”入口 | P1 |
| E-001 | 启发式追问 | 澄清意图不打断 | 无 | 输入含糊问题 | 返回1–2条引导式追问 | P1 |
| E-002 | 口语纠错要点 | 辅助信息不过载 | 开启“Speaking Tips” | 继续对话 | 每回合≤3条建议；不打断主对话 | P1 |
| E-003 | 要点笔记 | 一键生成可编辑卡片 | 无 | 点击“加入要点卡” | 生成可编辑卡片；可收藏/导出 | P1 |
| F-001 | 游客试用 | 免登录体验可完成 | 未登录 | 直接开始对话 | 至少 5 回合可完成；适时提示登录 | P0 |
| G-001 | 边界提示 | 合规不打断会话 | 触发敏感/越界问题 | 提出越界问题 | 以角色口吻返回边界说明+替代问法 | P0 |
| G-002 | 隐私透明 | 最小化存储可控 | 无 | 查看“隐私/记忆说明” | 说明本地存储与一键清空 | P0 |
| H-001 | 漏斗埋点 | 关键指标可观测 | 已接入埋点 | 从 LP 进入试用 | 记录 CTR/回合数/要点卡/D1/D7 等事件 | P0 |
| I-001 | TTS 播放 | 播放控制可用 | 已开启 TTS | 点击朗读→暂停/继续/重播 | 首包≤1.2s；播放流畅；控制有效 | P1 |
| I-002 | 失败回退 | 弱网/失败降级生效 | 弱网或TTS异常 | 触发播报失败 | 自动回退文本；提示稍后重试 | P1 |

3. 验收用例集

4. 首版通过标准（建议）

• 核心路径通过率 ≥97%；P95 首响应 ≤3.5s；P50 TTFB<2s、RTT<4s。

• 角色模板 ≥10；发现页 ≥12 角色卡；近 6–8 轮记忆有效。

• 3 分钟内达成 ≥5 回合的会话占比 ≥60%；要点卡使用率 ≥45%；D1 ≥25%，D7 ≥10%。

|  |  |
| --- | --- |
| 术语 | 定义 |
| TTFB | Time To First Byte，首字节时间。 |
| RTT | Round Trip Time，一问一答的整轮响应时延。 |
| LCP | Largest Contentful Paint，最大内容绘制。 |
| P50 / P95 | 中位数 / 95 分位的统计指标，用于衡量时延分布。 |

5. 术语表

**第 9 章｜ADR-001 模型选型**

我们选择 GPT OSS 120B 作为主模型，原因是其在角色扮演、结构输出、语言能力和稳定性上全面优于其它可用选项。回退与分流策略则使用 DeepSeek / Qwen 系列，保障性能与成本。

ADR-001｜模型选型（Primary Output Model）

【决策】

采用 GPT OSS 120B 作为主要输出模型（Primary），支持角色扮演、启发式追问、口语纠错要点、结构化输出。

【回退与分流】

回退模型：DeepSeek-V3 Thinking 或 Qwen3 32B/30B A3B Instruct

分流模型：Qwen、Doubao、DeepSeek 中的 30–34B 档

【对比范围】

Qwen系列、DeepSeek系列、Doubao系列、GLM 4.5、Kimi、MiniMax、Nemotron、xAI Grok、GPT-OSS系列

【为何选择 GPT OSS 120B】

1. 角色一致性最稳，边界控制好

2. 结构化输出（JSON/函数调用）合规率高

3. 长上下文能力强，适配技能/小剧本场景

4. 中英双语表现好，提示词友好

5. 平台托管/自托管可控，合规性强

6. 可通过策略分流平衡成本

【其他模型未作为主模原因】

DeepSeek 系列：推理快但结构化/人设一致性略逊

Qwen 系列：生态友好但合规率略低

Doubao 系列：语料友好但多轮对话稳定性不足

GLM / Kimi / MiniMax：短对话强，但复杂结构输出不稳

xAI / Llama3 / Nemotron：偏向代码或英文场景，不适合首版

|  |  |
| --- | --- |
| 指标 | 目标值 |
| P95 首 Token 延迟 | ≤ 1.8s |
| 服务端生成吞吐 | ≥ 20 tok/s |
| JSON 合规率 | ≥ 98% |
| 人设一致性 | ≥ 85/100 |
| 安全拒答准确率 | ≥ 95% |

【非功能性指标（NFR）】

【风险与缓解】

延迟/成本 → 批处理、预热、上下文裁剪、分流

结构化失败 → Schema 提示 + 重采样

成本风险 → 动态配额 + 路由优化

合规风险 → 审计、日志去标识、域内控制

【总结】

**第 10 章｜功能清单表（汇编版）**

• 归属 & 埋点：用于对齐 Owner 与可观测性，上线前需在数据字典定义事件与字段。

**功能清单表（Features List）**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 功能ID | 功能名称 | 用户价值 | 一句话用户故事 | 主要模块/流程 | 依赖 | 优先级 | 验收标准（90分制） | 归属 & 埋点 |
| F-001 | 文本对话（非语音） | 静音/无麦环境可用；稳定入口 | 作为用户，我希望直接打字与角色对话 | 输入框 → 上下文管理 → LLM 回复 → 展示 | LLM | P0 | 首屏回显≤100ms；响应成功率≥99%；多轮稳定性≥85/90 | 前端/后端｜埋点：会话开始/结束、轮次、耗时 |
| F-002 | 多轮上下文记忆 | 避免重复提问，提升沉浸 | 我希望角色能记住我刚刚说过的事 | 短期窗口记忆 + 轻量长期记忆（用户卡片） | KV 存储 | P0 | 最近10轮关键信息召回≥90/90；人格一致性≥80/90 | 后端｜埋点：记忆命中率、重提率 |
| F-003 | 角色背景知识库（人设/RAG） | 风格一致、世界观完整 | 我希望角色永远“像他自己”说话 | 人设模板→检索→提示词编排→应答 | KB/RAG | P0 | 语气一致性≥80/90；禁忌词召回率≥95% | 产品/后端｜埋点：人设命中、冲突告警 |
| F-004 | 小剧本生成（Hook/分镜骨架） | 快速给剧情钩子，降低创作门槛 | 我希望一键得到可续写的短剧情 | 模板选择→剧情生成→续写/改写 | LLM | P0 | ≥3类模板；续写连贯度≥80/90；生成≤3s | 后端｜埋点：模板使用率、续写率、导出率 |
| F-005 | 口语纠错（文本级） | 练口语但先用文本承载 | 我希望看到读错/写错处与改法 | 错误定位→纠错建议→例句/IPA | NLP/规则 | P0 | 定位准确性≥80/90；建议可读性≥80/90 | 后端｜埋点：纠错调用数、接受率 |
| F-006 | 对话摘要/结构化笔记 | 一键沉淀知识点与待办 | 我希望把长对话整理成要点 | 会话摘要→要点提取→TODO | LLM | P0 | 三档粒度；摘要可读性≥80/90；导出成功率≥99% | 前后端｜埋点：摘要使用率、导出率 |

**P0｜MVP（本期必须交付）**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 功能ID | 功能名称 | 用户价值 | 一句话用户故事 | 主要模块/流程 | 依赖 | 优先级 | 验收标准（90分制） | 归属 & 埋点 |
| F-101 | 情绪识别 & 情感回应（文本） | 被理解与被安抚 | 我希望角色能读懂我情绪并对应回应 | 文本情感分类→策略话术库→应答风格化 | 分类器 | P1 | 6类情绪识别 F1≥0.80；回应契合度≥80/90 | 后端/产品｜埋点：情绪分布、安抚成功率 |
| F-102 | 反诘树（Socratic） | 启发式学习、深度思考 | 我希望被高质量追问引导 | 问法树策略→分支追问→收束总结 | 策略引擎 | P1 | 三层连贯不兜圈；总结清晰度≥80/90 | 后端/产品｜埋点：追问层数、放弃率 |
| F-103 | 每日任务 & 打卡 | 形成习惯、提升留存 | 我希望每日收到小挑战并可打卡 | 任务派发→提醒→打卡→补记 | 定时器 | P1 | 通知抵达率≥98%；7日留存提升≥基准 X% | 前后端/运营｜埋点：完成率、连续天数 |
| F-104 | 线索板/关系图（Clue Board） | 可视化角色/事件关联 | 我希望把人物与线索拖拽整理 | 节点/边管理→画布交互→导出 | 前端图形库 | P1 | 100节点内流畅；PNG导出成功率≥99% | 前端｜埋点：节点数、导出率 |
| F-105 | 同人创作助手 | 提升二创效率与风格统一 | 我想快速写同人片段且不中断人设 | 风格固化→桥段补全→衔接检查 | LLM | P1 | 风格一致性≥75/90；桥段衔接≥80/90 | 后端/产品｜埋点：使用时长、完稿率 |
| F-106 | 可分享会话卡片 | 外部传播裂变 | 我希望把精选对话一键生成卡片 | 片段选择→排版模板→导出 | 前端渲染 | P1 | ≥3种模板；分享点击回流≥X% | 前端/增长｜埋点：分享率、回流率 |

**P1｜增强（次迭代）**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 功能ID | 功能名称 | 用户价值 | 一句话用户故事 | 主要模块/流程 | 依赖 | 优先级 | 验收标准（90分制） | 归属 & 埋点 |
| F-201 | 多人房间同聊 | 与朋友共创/同玩 | 我希望和朋友一起与角色对话 | 房间/成员管理→消息路由→并发仲裁 | 实时信令 | P2 | 3人并发无明显延迟；状态一致性≥99% | 前后端/RTC｜埋点：房间时长、掉线率 |
| F-202 | 角色联动/混战 | IP 碰撞、趣味提升 | 我希望多个角色同台互动不串台 | 多 persona 编排→轮替/联合应答 | Prompt 编排 | P2 | 串台率≤5%；用户评分≥80/90 | 后端/产品｜埋点：联动场次、满意度 |
| F-203 | 剧情任务/解谜（活动） | 深度沉浸、可复玩 | 我希望完成有分支的任务剧情 | 任务状态机→触发条件→结局收敛 | 状态机 | P2 | ≥3条可完成支线；重复游玩率≥X% | 后端/产品/运营｜埋点：完成率、分支流向 |
| F-204 | 成就系统/积分商城 | 激励与留存 | 我希望达成成就并兑换奖励 | 勋章/积分→兑换→风控 | 账户/风控 | P2 | 防刷通过；核算准确率≥99.9% | 后端/运营｜埋点：积分消耗、留存提升 |
| F-205 | A/B 话术实验平台 | 话术持续优化 | 我希望验证不同话术的效果 | 版本管理→用户分流→统计分析 | 实验平台 | P2 | 指标差异统计显著（p<0.05） | 后端/数据｜埋点：实验曝光/转化 |
| F-206 | 社区话题/活动日历 | 运营阵地 | 我希望报名活动并查看主题内容 | CMS→话题聚合→报名/提醒 | CMS | P2 | 报名成功率≥99%；内容发布≤5min 生效 | 前端/运营｜埋点：话题互动、报名转化 |

**P2｜拓展（社交/玩法）**

|  |  |  |  |
| --- | --- | --- | --- |
| 里程碑 | 时间窗口 | 目标功能 | 验收口径 |
| M1 | 本期 | F-001/002/003/004/005/006 | MVP 场景完整跑通，核心指标达标 |
| M2 | +1 迭代 | F-101/102/103/104/106 | 留存与分享能力上线 |
| M3 | +2 迭代 | F-105 + P1 打磨 | 二创与可视化体验提升 |
| M4 | +3 迭代 | F-201/202/203 | 社交与活动玩法形成闭环 |

**里程碑映射（简表）**

**字段补充说明**

• 主要模块/流程：在 PRD 的流程章节对应到具体时序图/状态机（本章保持摘要级别）。

• 依赖：仅列关键外部能力（如 LLM、RAG、实时信令、CMS）。

• 验收标准（90 分制）：结合性能/成功率等定量指标与人工质评（可读性/一致性）。